The Block Relevance (BR) analysis to aid medicinal chemists to determine and interpret lipophilicity

Giulia Caron, Maura Vallaro and Giuseppe Ermondi
The Block Relevance (BR) analysis to aid medicinal chemists to determine and interpret lipophilicity
Med. Chem. Commun., 2013,4, 1376-1381
DOI: 10.1039/C3MD00140G

The Block Relevance (BR) analysis to aid medicinal chemists to determine and interpret lipophilicity

Giulia Caron,[a] Maura Vallaro[a] and Giuseppe Ermondi*[a]

A major issue related to chromatographic determination of lipophilicity is about the conversion to log P. The interconversion of lipophilicity indexes can be made only if two systems express the same balance of intermolecular solute/system forces. The deconvolution of intermolecular interactions is generally obtained by solvation parameter models. Block Relevance (BR) analysis is a new tool specifically designed for medicinal chemists to interpret partitioning/retention phenomena in a very practical way. This paper describes the application of BR analysis to literature data (ElogP) and experimentally determined chromatographic indexes on a Supelcosil ™ LC-ABZ column for a series of 36 drugs. Results indicate that BR analysis is a solid and reliable tool that captures the main information encoded in any lipophilicity descriptor.

Introduction

The importance of lipophilicity in medicinal chemistry is widely known [1] in particular, to improve the efficiency of compounds development by minimizing the attrition rate and shortening the development time; and to satisfy the modern medicinal chemistry demands for physicochemical property determinations (e.g. lipophilicity) required for the early prediction of ADME profile [2-6].

As a result, there is a new interest in overcoming limitations in the determination and interpretation of lipophilicity data. Several methods are available to measure log P/log D. Nowadays most of the methods are automated, but many are still limited by solubility issues and detection limits. Because of these limitations, various chromatographic systems are very much used in pharma labs to determine lipophilicity [3,7-9]

A major issue related to chromatographic determination of lipophilicity is about conversion to log P[3,4]. In practice, if two systems express the same balance of intermolecular solute/system forces the correspondent lipophilicity indexes can be interconverted [2,3]. It was proven, with a few exceptions, that systems based on RP chromatography employing chemically bonded phases are unsuitable for estimating log P for compounds of diverse structure [3]. The reason being that the fundamental properties responsible for chromatographic retention tend to be different to those responsible for partitioning between octanol and water, especially the contribution from hydrogen bonding (HB) interactions.

The deconvolution of the fundamental intermolecular interactions responsible for chromatographic retention and partitioning is generally associated with the solvation models based on Abraham's solvation parameters [10,11] or similar descriptors. In short, Abraham's equations use multilinear regressions to relate the lipophilicity indexes (e.g. log P, log k) to solvation descriptors (e.g. V_x, R_2, $\Sigma\pi^H_2$, $\Sigma\alpha^H_2$, $\Sigma\beta^H_2$; see definitions in Annex S1, Supporting Information)[12].

Despite their great relevance in the comprehension of physicochemical events, solvation parameters suffer from a number of drawbacks[3] that limit their application, especially in HT environments. In particular it is very difficult to produce reliable descriptors (both experimental and calculated) for the complex molecular structures present in proprietary libraries.

VolSurf+ (VS+) is a computational procedure designed to describe and explore the physicochemical property space of a molecule starting from 3D interaction energy maps calculated with GRID force-field [13,14,15]. The basic concept of VS+ is to compress the information present in 3D maps into a few quantitative numerical descriptors (82 are used most often) that are very simple to understand [16]. Because of their non-experimental origin, only the chemical structure of the compound is necessary to obtain VS+ descriptors, which are mainly applied to model ADME-Tox events [17]. We already used these to model log P[18] . Since VS+ descriptors are intercorrelated, the derived statistical models are often obtained with Partial Least Squares regression algorithms (PLS).

Despite their statistical robustness, PLS models cannot be easily interpreted in mechanistic terms by medicinal chemists. To overcome this limit, some authors used VIPs plots [19].

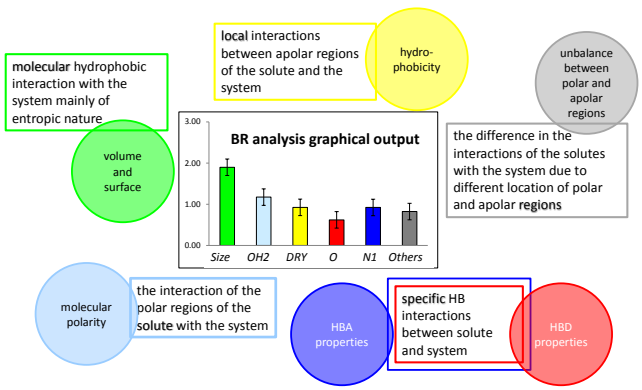

Fig. 1. BR analysis: interpretation scheme. A typical graphical output is shown in the center. The solutes' properties described by the blocks are in circular frames. The main interactions in which any block is involved are in rectangular boxes. A color code is always associated to any block (green for Size, cyan for OH2 (Water), yellow for DRY, blue for N1, red for O and grey for Others).

Recently we described the Block Relevance (BR) analysis, a new tool that facilitates the mechanistic interpretation of Quantitative Structure-(Chromatographic) retention Relationships (QSSR) based on PLS models and VS+ descriptors [20]. Briefly, the BR strategy groups the 82 VS+ descriptors in six easy-to-interpret blocks and graphically shows (Figure 1) the relevance of a certain block in the PLS model: the higher the value, the more important the block.

The organization of the VS+ descriptors in blocks (blocks definition and composition are given in Table S2 and S3, respectively, in the Supporting Information) enables a straightforward understanding of the investigated phenomena (e.g. chromatographic retention, partitioning) because blocks furnish an easy mechanistic interpretation based on the nature of the interaction of the solute with the environment represented by some tailored probes defined by the GRID methodology [13-15]. BR analysis is therefore an easy way to compare different systems.

The main goal of the paper is to describe the upgraded version of the BR analysis and, through it, shed some light on chromatographic indexes that can replace log P_{oct}. To reach this aim we firstly applied the new BR analysis to the 36 drugs published by Lombardo et al.[2] to verify the use of ElogP (chromatographic descriptor) to determine log P_{oct} (measured by shake-flask). Then for the same series of molecules we measured a panel of chromatographic indexes using different mobile phases and the same RP stationary phase (Supelcosil LC-ABZ column). BR analysis provided a rational approach for identifying rules to interconvert experimental lipophilicity indexes on the basis of the balance of the intermolecular interactions they express.

Results and discussion

The use of ElogP for the determination of log P to validate the BR analysis

In 2000 Lombardo and coworkers published the paper about ElogP, an easy, fast and accurate RP-HPLC method for the determination of log P_{oct} for 36 neutral drugs covering a log P range from -0.55 to 5.40 [2] (Table S1 in Supporting Information). The paper proved that ElogP expresses the same balance of intermolecular forces as the shake-flask log P data (difficult and long to obtain). To do that the authors used solvatochromic equations[2].

Here we decided to apply BR analysis [20] to the same set of data to validate our tool. BR analysis was therefore applied to a PLS model calculated as described below. ElogP and log P values were imported into VS+ as dependent variables (Y) and a relation between Y and VS+ descriptors (X) was sought using the standard PLS algorithm implemented in the software. Models validation was obtained through internal validation. Results are shown in Table 1.

Table 1. Final PLS models (n = number of observations, R^2 = cumulative determination coefficient, Q^2 = cross-validated correlation coefficient, LV = number of latent variables, RMSE = root mean square of the errors).

Model	n	R^2	Q^2	LV	RMSE	Data
log P	36	0.826	0.612	3	0.998	Lombardo et al. [2]
ElogP	36	0.806	0.635	3	0.957	Lombardo et al. [2]
log k'$_w$	36	0.836	0.716	3	0.656	Presented here
log k'80	28	0.865	0.583	3	0.255	Presented here
log k'70	28	0.871	0.666	3	0.277	Presented here
log k'60	35	0.739	0.522	3	0.468	Presented here
log k'50	34	0.755	0.536	3	0.502	Presented here
log k'40	*29*	*0.737*	*0.357*	*3*	*0.549*	*Presented here*
log k'30	*24*	*0.718*	*0.180*	*3*	*0.649*	*Presented here*

We are aware that some researchers in the QSAR field support internal validation, whereas others consider that internal validation is not a sufficient test to check the robustness of models, and external validation is necessary [21,22]. In this case, since the sample size is small, and thus holding a portion of it back for testing would be wasteful, it was preferred to use cross-validation (CV), with multiple rounds using different partitions (Table 1 shows Q^2 for the LOO procedure, results were similar with different partitions). The PLS models in Table 1 showed R^2>0.6 and Q^2>0.5, and satisfied the Tropsha et al. validation rules. [23] To further validate PLS models we also randomized the order of Y values which produced unacceptable R^2 and Q^2 values (data not shown).

The PLS outputs (VIPs and PLS coefficients) were then submitted to BR analysis to evaluate the relevance of the six blocks of descriptors to the model. Since in the original paper 5 out of the 82 VS+ descriptors were inconsistently assigned to a certain block, a few changes are introduced here. The final composition of the six blocks (definition and color codes are given in Figure 1) is reported in Table S2 of the Supporting Information.

BR graphical results for the Lombardo's dataset are in Figure 2 which shows a pair of plots for any lipophilicity index (log P and ElogP). The first indicates the relevance of the blocks to the model (Figures 1A and 1C for log P and ElogP, respectively). The second plot (Figures 1B and 1D for log P and ElogP, respectively) splits the contribution of any block in positive (BR (+)) and negative (BR (-)) portions. BR (+) indicates how much the considered block favors solutes partitioning in the octanol

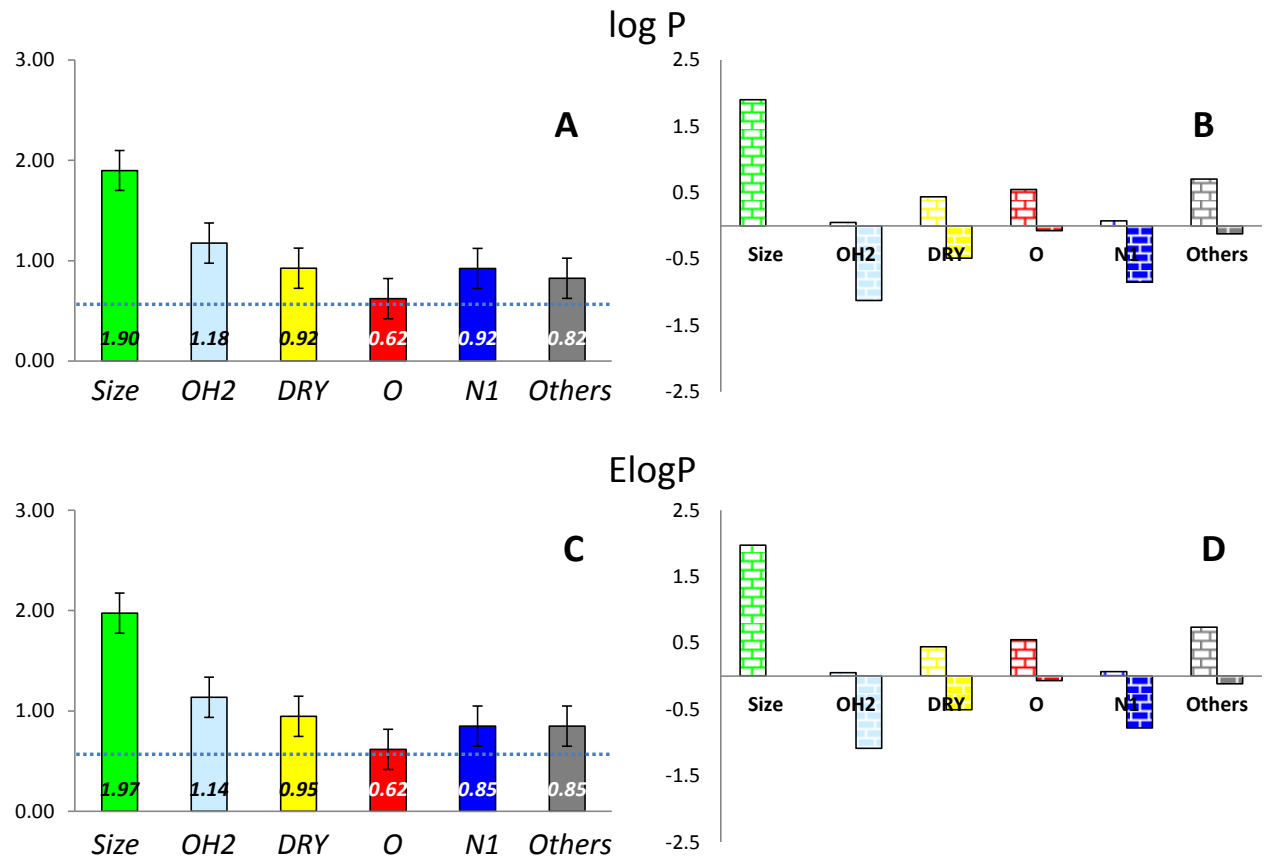

Fig.2 . BR analysis graphical outputs. The plot on the left (A and C) indicated the relevance of the block to the model (log P and Elog P). The second plot (B and D) splits the contribution of any block in positive (BR (+)) and negative (BR (-)) portions. Error bars obtained as described in the text are also reported. The dotted line shows the blocks significance threshold.

(stationary) phase whereas BR (-) indicates how much the considered block favors solutes partitioning in the aqueous phase. A complete BR analysis includes inspection of both plots.

Figure 2 shows that log P and ElogP show the same balance of intermolecular forces since blocks have similar relevance in the models. This finding supports the strategy of determining log P with the ElogP method and shows the reliability of BR analysis to obtain this information.

Please note that the original version of BR analysis did not include the uncertainty associated to any block. It is essential to compare different PLS models and to verify systems similarity. More details about this point are reported in the Supporting Information Figure S1.

Figure 2 outlines that the *Size* is the most important block (the significance and the main interactions associated to any block are schematized in Figure 1.). This finding indicates that as expected the larger the molecule, the higher the log P_{oct}. The *DRY* block (yellow) is partly positive and partly negative. The positive term could be due to hydrophobic interactions between the probe and apolar regions of the molecules. The negative contribution could be related to the interactions occurring between the solute and the aqueous phase of the system due to dipolarity/polarizability properties. These interactions in first approximation could be related to Abraham's π^{H}_{2} solute's properties (see Annex 1 in Supporting Information) and contribute to push solutes into the aqueous phase.

The *OH2* also named *Water* block is relevant in the opposite direction of the *Size* block. That means that the higher the ability of the solute to interact with water, the higher its propensity to have low log P_{oct}. (please refer to Figure 1 and Annex 1 of the Supporting Information for clarifying the difference between *Water* and *N1* and *O* blocks).

Another finding concerns with the hydrogen bond acceptor properties of the solutes (in blue, roughly superposable with Abraham's $\Sigma\beta^{H}_{2}$): the higher the HBA, the lower the log P. Conversely, the hydrogen bond donor properties of the solutes (in red, roughly superimposable with Abraham's $\Sigma\alpha^{H}_{2}$) have poor relevance here (slightly above the dotted line).

Interestingly, all these findings are consistent with Abraham's equations for log P_{oct} [2] (more details about the relationship between blocks and Abraham's parameters are given in Annex S1 of the Supporting Information).

The *Others* block is the main innovation of the BR analysis since it is related to the 3D structure of solutes and thus no corresponding term could be found in any solvation equation. The significance of the block is related to the relative location of the polar (apolar) regions. The positive sign of the block in the models takes into account the observation that closely located polar (apolar) regions partially mask their polarity (hydrophobicity) [24]. In Figure 3 DRDRAC was selected as an

example of the *Others* descriptors since in the final PLS models (Table 1) it is one of the most important (data not shown). Figure 3 graphically shows the DRDRAC descriptor for Bifonazole and Clotrimazole as the area of the triangle formed by the three atoms DR, DR and AC (in violet). These two compounds share three phenyl rings and an imidazole ring. As a result of the different arrangement of these features DRDRAC is larger for Bifonazole (32.8Å) than for Clotrimazole (20.0Å). In other words, the phenyl rings in Clotrimazole cannot fully express their hydrophobicity and this evidence is included in the final model through the DRDRAC descriptor.

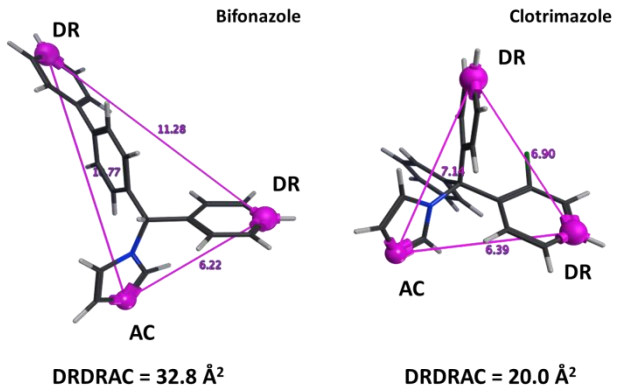

Fig. 3. DRDRAC descriptor is represented for two compounds of the dataset, Bifonazole (on the left) and Clotrimazole (on the right). The vertexes of the triangle of maximum area are shown in magenta. Shortly, all the atoms of a structure are classified by VS+ as Dry (DR), H-bond donor (DO) and H-bond acceptor (AC); then all possible triangles made from triplets of atoms are generated and the corresponding area calculated. Finally the triangle (DRDRAC here) with the maximum area is retained and its area is the descriptor used in VS+.

Analysis of lipophilicity indexes measured on a Supelcosil LC-ABZ column

The aim of the second part of the study is to shed light on the balance of the intermolecular interactions governing isocratic log k's and extrapolated log k'$_w$ values. This information is crucial for establishing if log k'$_w$ need to be obtained (longer and more difficult to obtain than log k' alone) to predict log P for a given system.

For the same series of 36 drugs discussed above we measured a series of lipophilicity indexes (log k'80, log k'70, log k'60, log k'50, log k'40, log k'30 and' log k'$_w$ where log k' is the logarithm of the capacity factor k' measured at a given organic solvent concentration, whereas log k'$_w$ is the logarithm of the capacity factor extrapolated to a 0% concentration of the organic solvent) with a chromatographic system already described in the literature [25]. Supelcosil LC-ABZ column is one of the most plausible candidates as a model for the octanol–water partition system[25]. The stationary phase is an alkylamidesiloxane-bonded, silica-based stationary phase with an embedded polar (amide) group close to the silica surface. [25] The system is, practically, the same as in ElogP apart from the presence of octanol.

Experimental data (log k') are reported in Table S1 in the Supporting Information. The log k'$_w$ values were calculated by extrapolation using linear regression to 100% water. Linear trends were verified for all compounds (some examples are shown in Figure S2 of the Supporting Information). We are aware that linearity was not always verified in the literature for the investigated system using different datasets of compounds. [25]

The correlation matrix (Figure S3 in the Supporting Information) shows large correlation coefficients among lipophilicity indexes. This result does not provide evidence that either log k's or log k'$_w$ can be used for estimating log P. To do that the same balance of intermolecular interactions should be verified for the involved systems [3]., The procedure described above to obtain and analyze PLS models was applied to verify this. The statistics of the PLS models are shown in Table 1. The last two models (log k'40 and log k'30) were discarded because of poor statistics probably related to a limited number of available values. Statistically acceptable PLS models were submitted to BR analysis as described above.

BR graphical outputs are reported in Figure 3 and Figure S4 (Supporting Information) and clearly prove that the balance of intermolecular forces varies with the amount of methanol in the mobile phase. In particular, the higher the amount of methanol, the lower the relevance of the size block to the model. This finding is in line with the observations by Pagliara et al. [26] and may be due to the more hydrophobic character and lower hydrogen-bond capacity of methanol compared to water. Interestingly, the variation in methanol content only poorly influences the contribution of the remaining blocks to the model, being the water block the most sensible out of the 5 remaining blocks.

BR analysis (Figure 4) shows that log k'60 determined on a Supelcosil LC-ABZ column with methanol–water mobile phases is a suitable model for estimating log P of neutral compounds as ElogP. Conversely, the extrapolated value (log k'$_w$) overestimates the contribution of the *Size* block and thus should not be used to obtain log P. This could also be related to the extrapolation process made with isocratic values obtained with high amount of MeOH in the mobile phase.

Taken together these findings are of paramount relevance since the identification of the "best" log k to mimic log P is an open question in the literature where many controversial opinions are reported [27]. In particular, the chromatographic descriptors that cannot replace log P could be good indexes per se, as already discussed [4] , and their potential should be highlighted by ad hoc informative tools as BR analysis.

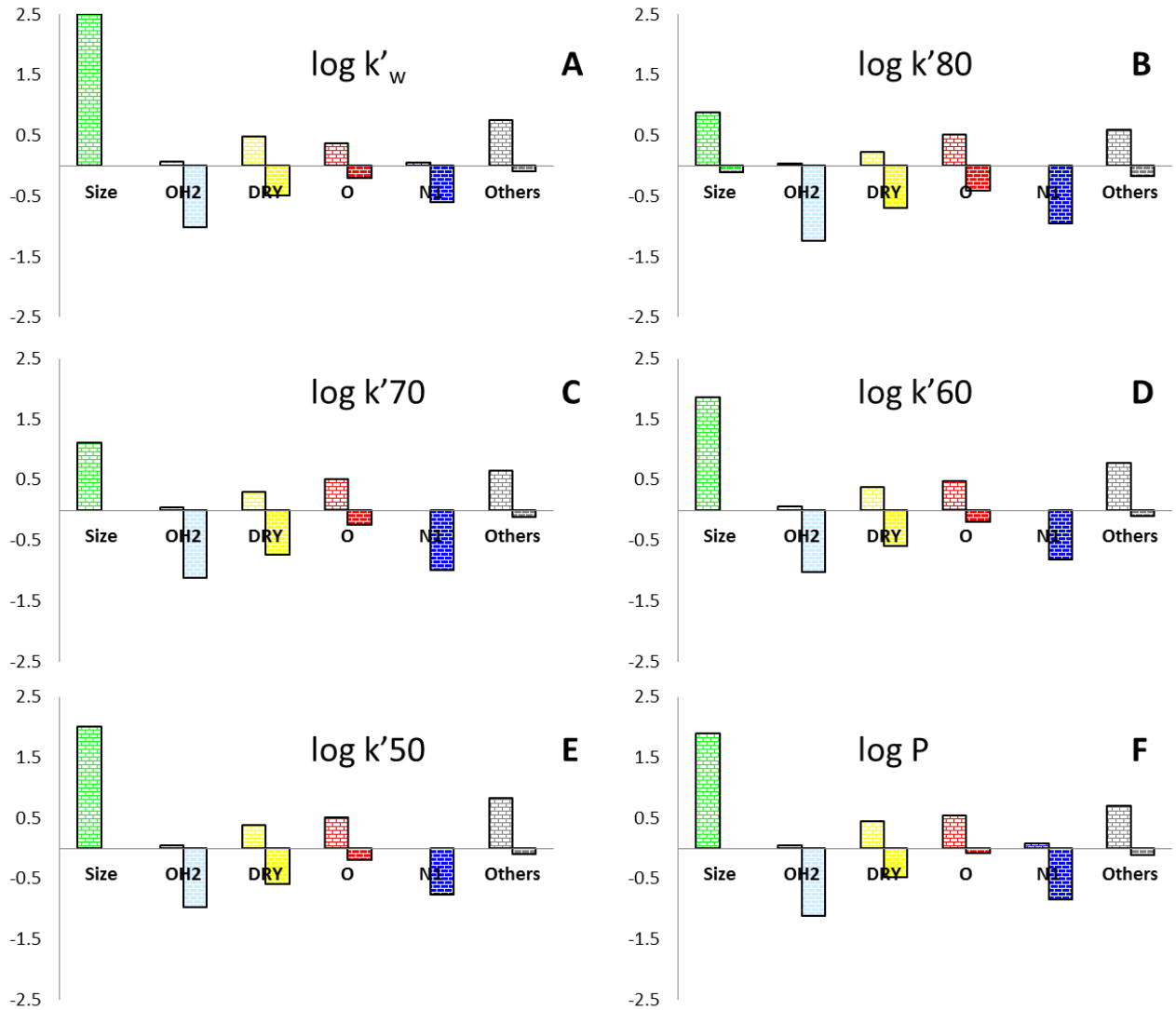

Fig. 4. BR plots with sign calculated for a series of log k' values determined using the same column with different mobile phase composition (see text for details)

Conclusions

The Block Relevance (BR) analysis is a new tool that facilitates the mechanistic interpretation of Quantitative Structure-(Chromatographic) Retention Relationships (QSSR). It is based on PLS models and VS+ descriptors.

With BR analysis it is possible to establish the main intermolecular forces governing a given physicochemical system and thus check if a given chromatographic descriptor can be used as a surrogate of log P. The application of BR analysis should, therefore, limit coarse estimations of log P from not checked chromatographic indexes.

BR analysis can replace solvation equations and go beyond them since the nature of VS+ descriptors introduces the 3rd dimension in the model. Moreover, these descriptors do not have any chemical space limitation and can be calculated for small molecules and peptides as well.

Finally, BR analysis adopts a convenient graphical format for delivering outputs specifically designed for a medicinal chemistry audience.

The application of BR analysis will be extended to more physicochemical systems,mainly to ADME-Tox topics, that often are described by the models that show good predictive abilities, but are poorly interpretable in mechanistic terms. This effort is expected to increase the confidence of medicinal chemists in QSAR models and thus to improve their practical daily use.

Methods

Materials and Methods

The dataset is represented by 36 drugs reported by Lombardo et al. [2] Most of the solutes were purchased directly from commercial sources (Sigma-Aldrich and Alfa Aesar). Benzodiazepines were purchased as dosage forms (aqueous solutions). All drugs were

used as received, in all cases. Deionized water and HPLC grade methanol (VWR) were used throughout.

The mobile phase consisted of 20 mM MOPS buffer at pH 7.4 and methanol in varying proportions from 30 to 80% v/v. For all mobile phases, the given pH is the pH of the buffer before the addition of organic modifier.

Samples were dissolved in methanol in a concentration range 50-100 μg/mL. The flow rate was 1 mL/min.

Injections of pure methanol were used to determine t_0, i.e., the dead time.

The retention time (t_R) were measured on a Supelcosil ™ LC-ABZ alkylamide column (Supelco, 5cmx4.6mm, 5μm packing, 120Å pore size). This phase has a unique deactivation technology which provides excellent reversed-phase performance for basic compounds, as well as those that are acidic, polar neutral, and non-polar (the ODS phase is pretreated with an electrostatic coating to suppress free silanophilic groups).

Four to five isocratic log k (capacity factor $k = (t_R - t_0)/t_0$) values were measured. The log k'_w values were calculated by extrapolation of isocratic log k_s against the mobile phase composition using linear regression. In all cases, the square of the correlation coefficient was > 0.99. Some examples of log k'_w extrapolation are shown in the Supporting Information (Figure S2).

A HPLC Varian ProStar instrument equipped with an 410 autosampler, a PDA 335 LC Detector and Galaxie Chromatography Data System Version 1.9.302.952 was used.

PLS models

SMILES codes were submitted to VolSurf+ (version 1.0.4, http://www.moldiscovery.com) using default settings and four probes (OH2, DRY N1 and O probes that mimic, respectively, water, hydrophobic, hydrogen bond acceptor and hydrogen bond donor interaction of the compounds with the environment).

PLS tools implemented in VolSurf+ were used.

BR analysis

BR analysis was performed as described elsewhere [20] .

Acknowledgements

This work has been supported by Ateneo Compagnia di San Paolo-2012-Call 2, LIMPET project

Notes and references

[a] *CASSMedChem research group, Dipartimento di Biotecnologie Molecolari e Scienze per la Salute at the Centre for Innovation, Università degli Studi di Torino, via Quarello 15, 10135 Torino, Italy. E-mail: giuseppe.ermondi@unito.it*

† Electronic Supplementary Information (ESI) available: [data used in the paper, blocks definition, the final composition of the six blocks, the variation of blocks values *vs* the number of LVs of the PLS model, examples of HPLC results, correlation matrix of lipophilicity indexes, A deeper insight in blocks significance]. See DOI: 10.1039/b000000x/

1. J. Arnott and S. Planey, *Expert Opin. Drug Dis.*, 2012, 7, 863.
2. F. Lombardo, M. Shalaeva, K. Tupper, F. Gao, M. Abraham, *J. Med. Chem.*, 2000, 43, 2922.
3. S. K. Poole and C.F. Poole, *J. Chromatogr. B*, 2003, 797, 3.
4. K. Valko, *J. Chromatogr. A*, 2004, 1037, 299.
5. G. Caron, and G. Ermondi, 2008 Lipophilicity: Chemical Nature and Biological Relevance, in Molecular Drug Properties - Measurement and Prediction (ed R. Mannhold), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527621286.ch12.
6. M. Shalaeva, G. Caron, Y. Abramov, T. O'Connell, M. Plummer, G. Yalamanchi, K. Farley, G. Goetz, L. Philippe, M. Shapiro, 2013, *submitted*.
7. S. Martel, D. Guillarme, Y. Henchoz, A. Galland, J.-L. Veuthey, S. Rudaz and P.-A. Carrupt, 2008 Chromatographic Approaches for Measuring Log P, in Molecular Drug Properties - Measurement and Prediction (ed R. Mannhold), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527621286.ch13.
8. A. Nasal, D. Siluk, and R. Kaliszan, *Curr. Med. Chem.*, 2003, 10, 381.
9. S. Gocan, G. Cimpan, and J. Comer, *Adv. Chromatogr.*, 2006, 44, 79.
10. M. Abraham, *Chem. Soc. Rev*., 1993, 22, 73.
11. M. Abraham, H. Chada, R. Leitao, R. Mitchell, W. Lambert, R. Kaliszan, A. Nasal, P. Haber,. *J. Chromatogr. A*, 1997, 766, 35.
12. M. Abraham, W. Acree, A. Leo, D. Hoekman, J. Cavanaugh. *J. Pharm. Sci.* 2010, 99, 2492.
13. P. Goodford, *J. Med. Chem.*, 1985, 28, 849.
14. D. Boobbyer, P. Goodford, P. Mcwhinnie, R. Wade, *J. Med. Chem.*, 1989, 32, 1083.
15. R. Wade and P. Goodford, *J. Med. Chem.*, 1993, 36, 148.
16. G. Cruciani, P. Crivori, P.-A. Carrupt, B. Testa. *J. Mol. Struct. (Theochem)*, 2000, 17–30.
17. P. Crivori, G. Cruciani, G., P. A. Carrupt, B. Testa. *J. Med. Chem*., 2000, 43, 2204.
18. G. Caron, and G. Ermondi, *J. Med. Chem*., 2005, 48, 3269.
19. G. Ermondi, G. Caron, M. Ravera, E. Gabano, S. Bianco, J. Platts, D. Osella, *Dalton Trans*., 2013, 42, 3482.
20. G. Ermondi and G. Caron, *J. Chromatogr. A*, 2012, 1252, 84.
21. S. Wold, M. Sjostrom, L. Eriksson, *Chemometr. Intell Lab.* 2001, 58, 109.
22. T. Scior, J. L. Medina-Franco, Q.-T. Do, K. Martínez-Mayorga, J. a Yunes Rojas, and P. Bernard, *Curr. Med. Chem.*, 2009, 16, 4297.
23. A. Tropsha, P. Gramatica, and V. K. Gombar, *QSAR Comb. Sci.*, 2003, 22, 69.
24. G. Caron and G. Ermondi, *Mini Rev. Med. Chem.*, 2003, 3, 821–830.
25. N. Dias, N. Nawas, C. Poole. *Analyst*, 2003, 128, 427.
26. A. Pagliara, E. Khamis, A. Trinh, P.-A. Carrupt, R.-S. Tsai, B. Testa, *J. Liq. Chromatogr*., 1995, 18, 1721.
27. R. Kaliszan, *Chem. Rev.* 2007, 107, 3212.